2022.03.11

Skunkworks Project: Radegen Bio's Enzymatic Toolkit

Fernando Andrade, M.S.,
Director the Skunkworks Division, Founder and CEO of Radegen Bio.www.radegenbio.com

Introduction:

The Radegen Bio Forging toolkit consists of a full suite of nucleic acid manipulation factors. The kit includes several ligases, polymerases, restriction enzymes, and molecular reporter proteins. Here a description, characterization and expression constructs are provided to be used for wet lab R&D. DNA binding proteins have similar characteristics that allows Radegen to use a common His-tag based for protein purification. Based on my analysis of the literature, it seems that there is a preference for protein purification of DNA binding enzymes. Polymerase preparation techniques use heat based purification using different columns to first denature the majority of the *E. coli* proteome leaving only enzymes that do not denature in high heat environments. This is the reason why polymerases are derived from thermophilic archaea. This method is more costly because it produces waist since there will be a percentage of enzyme loss during heating, and after running the preparation through multiple columns. A simpler and proven method for purifying proteins is needed that facilitates design but also allows for easy purification. A paper was identified, searching for clues of original methods that may have been adopted for large scale purification. A purification tag was found that was published by Dabrowski, S. and J. Kur (1998) that demonstrates a method that improves on previous tag-based protein methods that includes a 43 amino acid N-terminal tag that results in a protein with 700,000 U per mg. The 43 aa tag improved unit yield by 600,000 units per mg over previously reported methods. An observation made after this find suggest that the his-tag may change the conformation of the peptide since the projection point may imbed the his-tag at least 1 residue. This method is employed by Radegen for DNA binding proteins, the 43 aa tagged is termed Theta+43 domain, a proprietary feature of Radegen Bio's toolkit that provides a competitive advantage since purification is more efficient and less costly. The tag has been shown to not effect activity of a sensitive protein this much is clear since the majority of sources avoid their use. Simply extending the tag sufficiently may be an excellent method for purifying the majority of soluble proteins.

Development of this toolkit allowed me to contemplate how a synthetic biology researchers or companies are to deal with creating work with intellectual property that by themselves are a monumental task but are derived from publicly available sources. Today's work involves creating novel systems based on 70 year's worth of work and there is a plethora of ideas that may have been overlooked or simply forgotten. Synthetic Biologist have tasked themselves to use the collective knowledge up to date to form new function out of nature. The applications for this branch of science is literally world saving and the legacy model of intellectual property hording by major biotech corporations has promoted an unethical culture in biotech today. We have all experienced the demise of a colleague after sharing an idea. Oftentimes were left desperate when we have that one in a million idea and don't know what to do with it and feed our families at the same time. The United States academic community has literally trained the world in the life science discipline and it requires a special linage to produce a scientist with a comprehensive education to be able to first identify an interesting topic or problem, indicating the logic required to process the idea to begin with. The world has not propose such venture and stated it to be impossible. Enzymatic DNA synthesis is possible and here I report on enzymes that can be used for this purpose.

Keeping in mind the nature of open source biology, it is important to acknowledge that The American Society for Microbiology, under the direction of Sam Kaplan brought microbiology to the world because the world needed it by focusing his efforts on disseminating information via ASM publications. Microbiologist trained in molecular genetics are perfectly suited for both developing and working in settings that produce completely sustainable synthetic DNA using *E. coli* as the production chassis, and the tools developed from the work of ASM members. Samuel Kaplan achieved in two key areas, he mad ASM what it is today by ensuring the dissemination of information for advancing the art, and developing a Microbiology and Molecular Genetics Department with all the correct areas of expertise from literal world leaders in their field. Sometimes ideas are forehead slapping and other times they take expert training in systematic approaches to addressing problems.

Here, I specifically present the application of rational design for engineering synthetic proteins. A description of each one is provided and its specific use. I am proud to announce that I have decided to protect this publication by CC BY-NC-SA 4.0 Creative Commons licensing architecture since improvements can be made over the concepts created here. Strategies and sequences not previously employed by intellectual property protection are free to be modified and shared under the same terms meaning that if you have intellectual property that is derived from this work, then you can modify with attribution for non-commercial use. In this case I do see, as anyone else would a few improvements that can be made. Industry can get licensing rights to use the novel enzyme described here and make modification under explicit permission since this license both protects the rights of a licensee and licensor. For example, after expiration of a license, one can then re-license a protected item to someone else. This structure guarantees that the work started under an open-source culture is always maintained since improvements are published with the same license, remain open source for non-commercial use but protects the rights of the licensor for exclusive use of the approved improvement. Commercial entities cannot begin an improvement process before negotiating an improvement and compensation over the course of the licensing terms. Legal term presented here is part of this creative work and this statement although indeed a mask for using this approach for intellectual rights protection. Litigation will be a an act of corporate extortion directly directed to the scientific community that first made the discoveries that most molecular microbiologist simply know like the back of their hand. The synthetic biology discipline, one that is currently almost exclusively staffed with academics are genuinely forming a new reality with concepts that result in wonderful work with commercial potential. Protection is required because it take decades of work to even know enough to identify that there is a problem or that a novelty is genuine. Any protest to works coming from corporate biotech against an open source company will be laughed at since support for the corporate giant, that relies on bully tactics to convince others that their ideas are not novel or to simply steal a laptop rely on breaking the law in every way imaginable. Some of the most brilliant minds have the potential of being in jeopardy for simply being brilliant.

This document provides a detailed description of the Theta+ molecular toolkit along with details and results of the *in silico* development of the toolkit. This document contains trade secrets and is protected under copyright and trademark protection.

The Theta+ Polymerase Suite: This set of enzymes is composed of one high fidelity and one low fidelity polymerases with improved processivity and efficiency. In a PCR reaction they both have fidelity rates similar to the native enzyme but have faster reaction rates and are able to polymerize larger DNA fragments. Both polymerases have processivity values of 0.98 (Microscopic processivity (PI)[a] and both extend primers by 55 nt per protein/DNA binding event (PI) (Average primer extension length (nt) [1/(1 ± PI)][a]). This feature allows PCR reactions to have the same thermocycling program regardless of the polymerase being used. This system also includes a terminal 5′ – 3′ polymerase used in *de novo* DNA forging. The Theta43 domain unifies all proteins described here into a common system.

1. **ThetaPfu+S Polymerase –** An enzyme based on an open-source Pfu polymerase first isolated from *Pyrococcus furiosus*. The enzyme is an archaeal chimera fusing Pfu with a thermo stable non-specific DNA binding peptide, Sso7 from *Sulfolobus solfataricus*. This fusion protein has the same fidelity as Pfu but with improved efficiency. ThetaPfu-S Pol. is a proprietary next-generation Pfu polymerase that is used across all production and R&D applications at Radegen Bio. a. **Error rate**: 2.8 x 10^{-6} (one error every 2.8 million bp)

b. **Notable features**: 5′ – 3′ polymerization; 3′ – 5′ exonuclease activity; max length: 15kb; efficiency: 10kb with 1 min extension time.
c. **Published reaction conditions for PCR:** PCR buffer contained 20 mM Tris±HCl pH 8.8, 10 mM (NH4)2SO4, 0.1% Triton-100, 2 mM MgCl2 and 200 mM each dNTPs with 10 mM KCl for Pfu and 60 mM KCl for Pfu-S. The cycling protocol was 95°C for 20 s; 20 cycles of 94°C for 5 s and 72°C for 30 s (A) or for 60 s (B) or for 2 min (C); 72°C for 7 min.
d. **Efficiency :** when 10 U/ml Pfu-S was used, the same 5 kb target can be ampli®ed with a 30 s/cycle extension time. When a 2 min/cycle extension time was used, products as long as 15 kb were clearly detected with Pfu-S. 1kb/8 sec

2. **ThetaTaq+S Polymerase** – An enzyme based an open source Taq polymerase first isolated from *Thermococcus aquaticus*. The enzyme is an archaeal chimera fusing Taq with a thermo stable non-specific DNA binding peptide, Sso7 from *Sulfolobus solfataricus*.This fusion protein has the same fidelity as Taq but with improved efficiency. ThetaTaq-S Pol is a proprietary next-generation Taq polymerase that is primarily used for analytical techniques like qPCR or running an agarose gel. ThetaTaq-S maintains A-tailing activity and is an important molecular tool for applications like TA cloning.

a. **Error rate**: 5.6×10^{-5} (one error every bp 560,000)
b. **Notable features**: 5′ – 3′ polymerization; 5′ – 3′ (exo-); 3′ – 5′ (exo-); max length: 5kb;
c. **Published reaction conditions for PCR**: The PCR buffer contained 10 mM Tris±HCl pH 8.8, 2 mM MgCl2, 200 mM each dNTPs and 0.1% Triton-100 with 10 mM KCl for Taq(D289) and 50 mM KCl for S-Taq(D289) and Taq. The cycling protocol was: 95°C for 20 s; 20 cycles of 94°C for 5 s and 72°C for 30 s (A) or for 60 s (B) or for 2 min (C); 72°C for 7 min.
d. **Efficiency:** 20 U/ml enzyme and a 1 min/cycle extension time amplified a 5 kb target; 1kb/12 sec

3. **ThetaDtd+ Polymerase** – An enzyme based on a engineered dTd from terminal deoxynucleotidyl transferase from *Zonotrichia albicollis*, TdtR335L-K337G. This peptide is further engineered with the addition of the Theta43 tag and C to A mutations in all but 3 cystine residues. This is a completely novel proprietary DNA polymerase used in Radegen Bio's *de novo* DNA forging system. This enzyme is amenable for both Dtd-dNTP conjugate de novo synthesis and Dtd de novo synthesis using dNTPs protected on the 3′ OH group by a ONH2 group. 5 versions of the polymerase were developed to include 2 ThetaDtd+ fused with SsoT and 2 ThetaDtd+ fused with Ssod7. All 4 enzymes will be tested for improved de novo strand elongation in terms of processivity and efficiency using both dNTP protective strategies.

a. Radegen Bio ssDNA synthesis platform is a revolutionarily simple and robust DNA synthesis platform that uses a genetically engineered terminal deoxynucleotidyl transferase from *Zonotrichia albicollis*, ThetaTdt+ and was wholly developed by the Radegen Bio Skunkworks division. This platform can generates dsDNA fragments >200 bp. The fragments produced by this process are used by Radegen Bio for the de novo construction of circular DNA preps. It's simplicity lies in the use of standard liquid handling robotics designed for preparing DNA preps using magnetic beads (dedicated custom design instrumentation is being investigated for feasibility and cost). The system is based on a 96 well format microtiter plate footprint since the processivity per plate is faster with smaller scales of individual reactions. Since a standard Eppendorf liquid handling robot is a fraction of the cost (< $10,000 USD) compared to all DNA synthesis instruments currently sold, this approach allows for scaling of yield by using multiple liquid handling units. Pipet tip consumption can be minimized by using tips multiple times. The synthesis reaction occurs on a priming ssDNA oligo substrate, termed ϴgattaca$_N$ and with a sequence identity of 5′ – (Gattaca)$_N$ – 3′ that is tethered to 1 micron streptavidin coated magnetic bead via a 5′ linked biotin group. The reaction substrate is added to the wells of a 96 well plate designed for silica-based DNA purification. The reaction can also be carried out in a 96 well plate with the use of a magnetic module in the robotic liquid handler. The reaction commences with the addition of a protected dNTP (3′-ONH2-dNTPs (Firebird Biomolecular Sciences, LLC, US) or a dNTP-Dtd conjugate; both protects the 3′ end of the elongating strand after incorporation of the dNTP and with ϴTdt. The reaction is incubated at 30 C for 10 min followed by a wash step to remove the polymerase and excess dNTP. The ONH2 or Dtd protective group is then removed by treatment with a sodium nitrite buffer followed by a wash step to remove deprotection buffer. These

steps are repeated until the desired DNA molecule is elaborated. After the polymerization steps are complete the magnetic beads are resuspended and removed from the reaction plate and transferred to a 96 well plate PCR plate. dsDNA libraries are made using a universal primer sequence that binds to a termina adaptor at the end of the ssDNA tethered to the magnetic beads and a primer binding site found on ⊖gattaca$_N$ priming tethered oligo. The dsDNA prep is then removed using magnetic bead purification. **This process is referred to as DNA forging since it differs from processes that use chemical synthesis. Chemical synthesis relies on the reactivity of a reactive functional group produced by the chemical conditions of the buffers used (such as pH or the presence of a chemical catalyst) in each elongation step. Enzymatic DNA synthesis is more akin to a forging process since the enzyme needs to make physical contact (like a hammer striking metal to forge a functional structure) to complete an extension step.** Thus, **⊖+ dna (Theta plus DNA) forging is a proprietary enzymatic *de novo* stepwise DNA synthesis system that produced dsDNA preps with fragment length > 200 bp and composed of a distinct desired sequence. The dsDNA fragment preps produced by the ⊖Radegen Bio Foundry are intended for use in the forging of circular DNA molecules. Theta+ *de novo* DNA Forging was conceived and developed by the CSO/CEO of Radegen Bio, Fernando Andrade, M.S**. and is described here for the first time, is proprietary and for the exclusive use by the ⊖Radegen Bio dsDNA Foundry. This system will not be sold nor described to the public aside from the bold red text above. The rational for calling Radegen Bio's DNA synthesis process "forging" can be disclosed to the public with the sentence above in bold blue text.

4. **ThetaPfUTPase** – An enzyme based on a open source dUTPase that enhances the Theta+ Archaeal Pol performance in terms of larger yields and maximum fragment size. This occurs by degrading dUTP, preventing Theta+ Archaeal Pol. from incorporating dUTP into an amplifying fragment. dUTP incorporation inhibits further dNTP incorporation and is problematic enough with Pfu based polymerase that the presence of dUTP has a drastic impact on DNA yield. Theta+ PfUTPase is a proprietary enzyme used in enzymatic cocktails for building DNA constructs > 15,000 bp.
5. **Theta+SsoT –** a ssDNA binding protein shown to enhance processivity of DNA polymerase. This peptide can be added to a PCR reaction to improve yield and processivity. The Theta+ polymerase suite has 2 Tdt-SsoT Fusion polymerases (proprietary) for use in dsDNA forging.
6. **Theta+TEV** – a protease used in protein purification process and complementary to the Theta43 Tag.
7. **Theta+Sso7d** – a dsDNA binding protein shown to enhance processivity of DNA polymerase. This peptide can either be fused to either the N or C terminus of a polymerase and has be shSown to increase efficiency in terms of yield and maximum fragment length.
8. **Theta43 domain** – A multipurpose purification tag that has been shown to maintain normal function of DNA binding proteins, specifically DNA polymerases. Comparison between this 43aa domain and a previously reported 12 aa tag showed that the shorter tag had an inhibitory effect while the 43 aa tag was benign. This tag provides a 6x His tag, a thrombin cleavage site, an S-tag, and a TEV site that produces a tag-less protein.
9. **Theta+ISO –** Based on an open source iso-thermal polymerase first isolated from *Bacillus stearothermophilus*. High fidelity Bst. It has an error rate of about 7.8x10^-7, a fidelity about 20 times higher than standard Bst, and better than Phusion in High-GC buffer. The specifically patented enzyme is the polymerase, which is a NEB proprietary strand displacing polymerase with high fidelity. Luckily Bst-HF is precisely that, a high-fidelity strand displacing polymerase. The kit is supposed to run at 50°C (-60°C), which also happens to be within the optimal temperature range for Bst. One other use for Bst-HF, or even better fusion variants on a Taq scaffold of that enzyme, would be for high-GC sequences. The addition of a bit of thermotolerant strand displacing polymerase lets one amplify DNA sequences with well over 80% GC with ease, and boosts amplification of large fragments due to its' ability to overcome complex/problematic sequences
10. **Theta+EXO –** Based on a 5' – 3' exonuclease first isolated from T5 bacteriophage.
11. **Theta+LGT –** Based on a thermo stable ligase first isolated from *Thermophilus aquaticus*.

The data provided below provides data performance improvements that result from adding a DNA binding peptide to PCR.

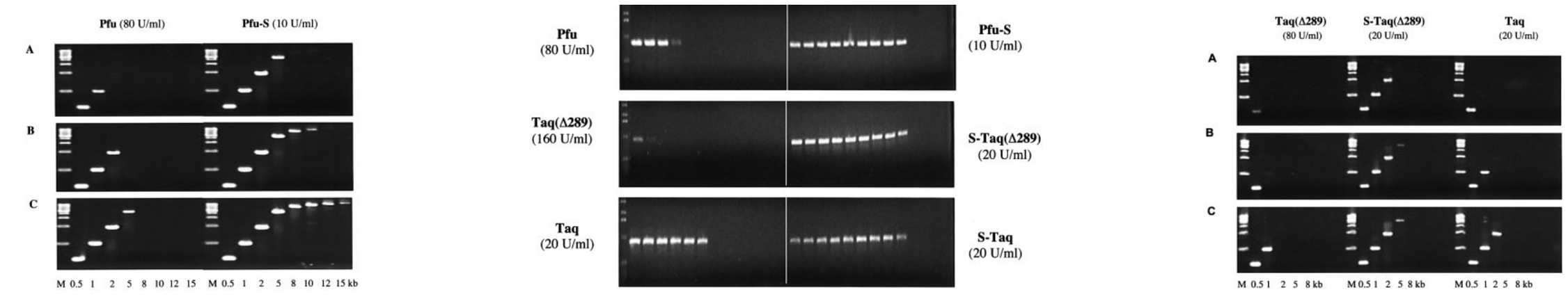

Figure 1. Agarose gel electrophoresis analysis of PCR efficiencies for Pfu, Pfu-S, Taq, Taq(Δ289), Taq-S and Taq(Δ289)-S. A.Comparison between Pfu and Pfu-S of total PCR product generated as a product of extension time and template length. Pfu-S amplifies DNA templates > than 5kb under conditions where extension times are greater than 60 seconds. Wild-type Pfu does not exhibit amplification of targets greater than 5 kb under 2 min and 7 min extension length conditions. B. Agarose gel electrophoresis analysis determining salt tolerance of the tested polymerase variants. Mutants augmented with the Ssod7 domain had an increased tolerance to KCl up to 120 mM and concentrations above this threshold resulted in polymerization abolishment instead of a gradient decrease. A.Comparison between Taq, Taq(Δ289), Taq-S and Taq(Δ289)-s of total PCR product generated as a product of extension time and template length. Taq(Δ289)-S amplifies DNA templates < or = to 5kb under conditions where extension times are greater than 60 seconds. Wild-type Tan does not exhibit amplification of targets greater than 2 kb under 2 min and 7 min extension length conditions.Taq(Δ289) experienced a general loss of function since there was a loss of fragment size capacity under all conditions, yielding fragments no greater than 1kb regardless of incubation time

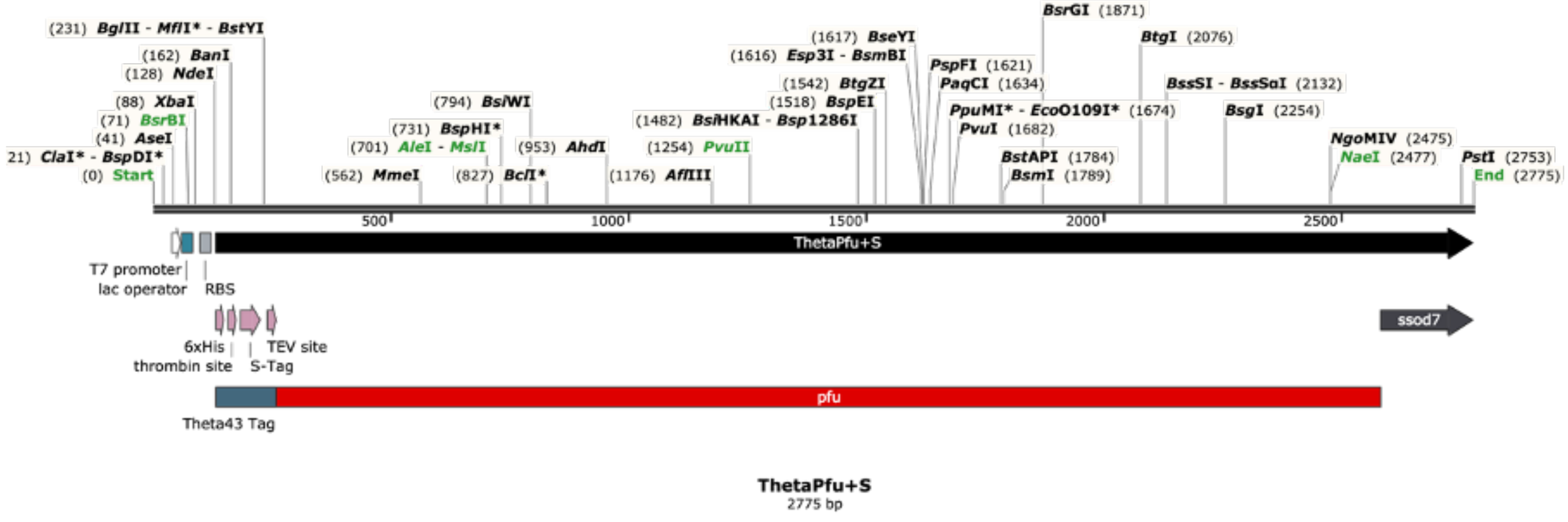

Figure 2. A cartoon representation of the ThetaPfu+S expression construct and protein domains. The Pfu+S is a novel 881aa protein containing the Theta43 domain, a polymerase domain from Pfu and a DNA binding domain, ssod7.

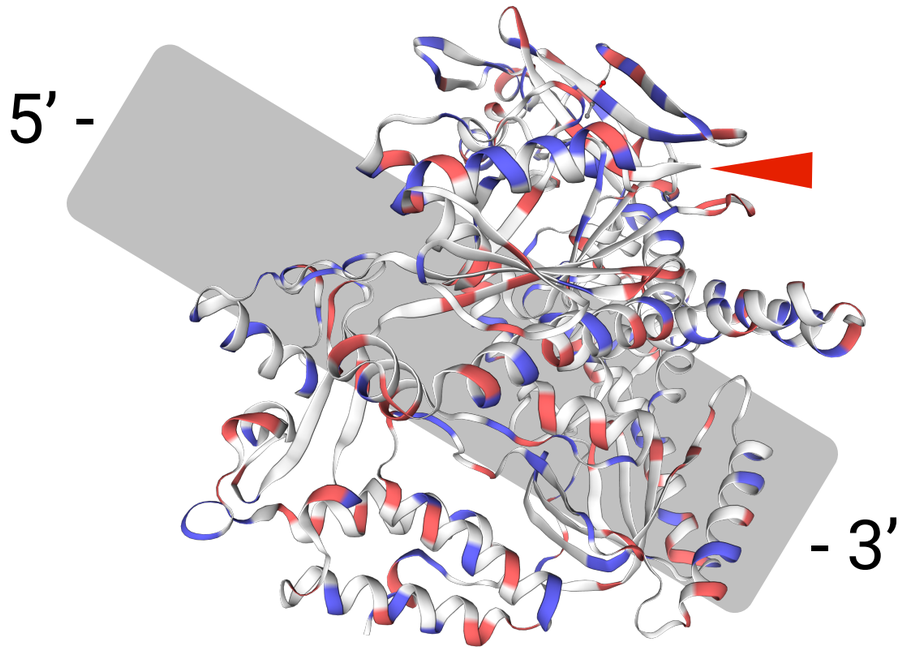

Figure 3. 3D model of Pfu showing the native N terminus of the protein and highlighted by a red triangle with the elongated tip pointing at the last residue on the N termini. The N terminus is directed away and towards the surface from the catalytic core. The grey rectangle depicts DNA and is associated to the enzyme structure to illustrate that the catalytic core is on the opposite site of N terminus outward projection point on the top exterior of the enzyme structure. Models for Dtd polymerases did not contain enough N or C terminal residues to make a prediction regarding fusion location relative to the native protein.

In silico development of Theta+ Polymerase suite.

1) ThetaDtd+

The ThetaDtd+ polymerase is a synthetic protein developed from an engineered Dtd polymerase. This enzyme was chosen with the specific aim of finding a Dtd enzyme that was more efficient since a strategy that produces a sustainable method for deprotecting after an elongation step. A protein chaise with improved catalytic rates but with a deprotection methods that left a molecular scar but did not generally interfere with catalysis since a reaction cycle is only 30 seconds. Chemical removal of the protective group on the polymerized strand relies on suppliers. Radegen Bio is committed to sustainable solutions and as company that uses the tools of the molecular biology revolutions, proteases are a preferred method for removing the polymerase from bead bound DNA. This enzyme is versatile since both dntp-polymerase conjugates’ or a chemically protected by an o-allyl bond, they can both be removed by UV light or chemicals, making it a lithographic enzyme that is versatile These enzymes were right under our nose since Dtd are used for tagging the 5’ end of an oligo with biotin and sold by Sigma. This Dtd is unique in that it uses the Theta+43 tag in some iterations and combines a enzyme with improved function that resulted from improvements made over wildtype that improve hydrogen binding. Versions containing fusions with 2 archaeal DNA binding proteins, independently of course, to test for novel de facto ligase functionality.

```
> XM_026799623.1
MDRFKAPAVISQRKRQKGLHSPKLSCSYEIKFSNFVIFIMQRKMGLTRRMFLMELGRRKGFRVESELSDSVTHIVAENNSYLEVLDWLKGQAVGDSSRFELLDISWFTACMEAGRPVDSEVKYRLMEQSQSLPLNMPALEMPAFIATKVSQYSCQRKTTLNNYNKKFTDAFEVM
AENYEFKENEIFCLEFLRAASLLKSLPFSVTRMKDIQGLPCVGDQVRDIIEEIIEEGESSRVNEVLNDERYKAFKQFTSVFGVGVKTSEKWYRMGLRTVEEVKADKTLKLSKMQKAGLLYYEDLVSCVSKAEADAVSLIVKNTVCTFLPDALVTITGGFRRGKNIGHDIDFLIT
NPGPREDDELLHKVIDLWKKQGLLLYCDIIESTFVKEQLPSRKVDAMDHFQKCFAILKLYQPRVDNSTCNTSEQLEMAEVKDWKAIRVDLVITPFEQYPYALLGWTGSRQFGRDLRRYAAHERKMILDNHGLYDRRKRIFLKAGSEEEIFAHLGLDYVEPWERNA
>XP_036017403.1 DNA nucleotidylexotransferase isoform X2 [Mus musculus]
MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDLVLFILEKKMGTTRRAFLMELARRKGFRVENELSDS
VTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLIECMGAGKPVEMMGRHQLVVNRNSSPSPVPGSQ
NVPAPAVKKISQYACQRRTTLNNYNQLFTDALDILAENDELRENEGSCLAFMRASSVLKSLPFPITSMKD
TEGIPCLGDKVKSIIEGIIEDGESSEAKAVLNDERYKSFKLFTSVFGVGLKTAEKWFRMGFRTLSKIQSD
KSLRFTQMQKAGFLYYEDLVSCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMTGHDVDFLITS
PEATEDEEQQLLHKVTDFWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFLILKLDHGRVHSEKS
GQQEGKGWKAIRVDLVMCPYDRRAFALLGWTGSRFERDLRRYATHERKMMLDNHALYDRTKRVFLEAESE
EEIFAHLGLDYIEPWERNA

ClustalOmega Analysis

XM_026799623.1      MDRFKAPAVISQRKRQKGLHSPKLSCSYEIKFSNFVIFIMQRKMGLTRRMFLMELGRRKG    60
XP_036017403.1      MDPLQAVHLGPRKKRPRQLGTPVASTPYDIRFRDLVLFILEKKMGTTRRAFLMELARRKG    60
                    ** ::*  :  ::** : * :*  *  *:*:* ::*:**:::*** *** *****.****

XM_026799623.1      FRVESELSDSVTHIVAENNSYLEVLDWLKGQAVGDSSRFELLDISWFTACMEAGRPVDSE    120
XP_036017403.1      FRVENELSDSVTHIVAENNSGSDVLEWLQLQNIKASSELELLDISWLIECMGAGKPVEMM    120
****.***************  :**:**: * :  **.:*******:  ** **:**:

XM_026799623.1      VKYRLMEQSQSLPLNMPA-LEMPAFIATKVSQYSCQRKTTLNNYNKKFTDAFEVMAENYE    179
XP_036017403.1      GRHQLVVNRNSSPSPVPGSQNVPAPAVKKISQYACQRRTTLNNYNQLFTDALDILAENDE    180
                     :::*: : :* *  :*.  ::**  ..*:***:***:*******: ****:::*** *

XM_026799623.1      FKENEIFCLEFLRAASLLKSLPFSVTRMKDIQGLPCVGDQVRDIIEEIIEEGESSRVNEV    239
XP_036017403.1      LRENEGSCLAFMRASSVLKSLPFPITSMKDTEGIPCLGDKVKSIIEGIIEDGESSEAKAV    240
                    ::***  ** *:**:*:****** :* *** :*:**:**:*:.*** ***:****..: *

XM_026799623.1      LNDERYKAFKQFTSVFGVGVKTSEKWYRMGLRTVEEVKADKTLKLSKMQKAGLLYYEDLV    299
XP_036017403.1      LNDERYKSFKLFTSVFGVGLKTAEKWFRMGFRTLSKIQSDKSLRFTQMQKAGFLYYEDLV    300
                    *******:** ********:**:***:***:**:.::::**:*::::*****:*******

XM_026799623.1      SCVSKAEADAVSLIVKNTVCTFLPDALVTITGGFRRGKNIGHDIDFLITNPGPRED--DE    357
XP_036017403.1      SCVNRPEAEAVSMLVKEAVVTFLPDALVTMTGGFRRGKMTGHDVDFLITSPEATEDEEQQ    360
***.: **:***::**::* *********:********  ***:*****.*   **  ::

XM_026799623.1      LLHKVIDLWKKQGLLLYCDIIESTFVKEQLPSRKVDAMDHFQKCFAILKLYQPRVDNSTC    417
XP_036017403.1      LLHKVTDFWKQQGLLLYCDILESTFEKFKQPSRKVDALDHFQKCFLILKLDHGRVHSEK-    419
***** *:**:*********:**** * : *******:******* **** : **....

XM_026799623.1      NTSEQLEMAEVKDWKAIRVDLVITPFEQYPYALLGWTGSRQFGRDLRRYAAHERKMILDN    477
XP_036017403.1      -----SGQQEGKGWKAIRVDLVMCPYDRRAFALLGWTGSR-FERDLRRYATHERKMMLDN    473
                           * *.*********: *:::  :********* * *******:*****:***
XM_026799623.1      HGLYDRRKRIFLKAGSEEEIFAHLGLDYVEPWERNA 513
XP_036017403.1      HALYDRTKRVFLEAESEEEIFAHLGLDYIEPWERNA 509
                    *.**** **:**:* *************:*******
```

Mutations performed R335L-K337G
The only cysteine residues in the construct are two buried cysteines depicted (Cys155, Cys404) and Cys302 (yellow) that serves as attachment point for the linker.

> ThetaDtd+

```
MDRFKAPAVISQRKRQKGLHSPKLSASYEIKFSNFVIFIMQRKMGLTRRM     50
FLMELGRRKGFRVESELSDSVTHIVAENNSYLEVLDWLKGQAVGDSSRFE     100
LLDISWFTAAMEAGRPVDSEVKYRLMEQSQSLPLNMPALEMPAFIATKVS     150
QYSCQRKTTLNNYNKKFTDAFEVMAENYEFKENEIFALEFLRAASLLKSL     200
PFSVTRMKDIQGLPAVGDQVRDIIEEIIEEGESSRVNEVLNDERYKAFKQ     250
FTSVFGVGVKTSEKWYRMGLRTVEEVKADKTLKLSKMQKAGLLYYEDLVS     300
CVSKAEADAVSLIVKNTVATFLPDALVTITGGFRLGGNIGHDIDFLITNP     350
GPREDDELLHKVIDLWKKQGLLLYADIIESTFVKEQLPSRKVDAMDHFQK     400
CFAILKLYQPRVDNSTANTSEQLEMAEVKDWKAIRVDLVITPFEQYPYAL     450
LGWTGSRQFGRDLRRYAAHERKMILDNHGLYDRRKRIFLKAGSEEEIFAH     500
LGLDYVEPWERNA
```

Improved DNA:

```
ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAA     50
AGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTA     100
ACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATG     150
TTCCTGATGGAACTGGGTCGTCGTAAAGGTTTCCGTGTTGAATCTGAACT     200
GTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAG     250
TTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAA     300
CTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCGGT     350
TGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGC     400
TGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCT     450
CAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATT     500
CACCGACGCTTTCGAAGTTATGGCTGAAAACTACGAATTCAAAGAAAACG     550
AAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTG     600
CCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGG     650
TGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAATCTT     700
CTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAG     750
TTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCG     800
TATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAAC     850
TGTCTAAAATGCAGAAAGCTGGTCTGCTGTACTACGAAGACCTGGTTTCT     900
TGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACAC     950
CGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCC     1000
GTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACCAACCCG     1050
GGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAA     1100
AAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTA     1150
AAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAA     1200
TGCTTCGCTATCCTGAAACTGTACCAGCCGCGTGTTGACAACTCTACCGC     1250
TAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTA     1300
TCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTG     1350
CTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGTCGTTACGC     1400
TGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTC     1450
GTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCAC     1500
CTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCT
```

> ThetaDtD+

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtGAAAACCTGTACTTCCAG

ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAAAGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTAACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATGTTCCTGATGGAACTGGGTCGTCGT
AAAGGTTTCCGTGTTGAATCTGAACTGTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAGTTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAACTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCG
GTTGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGCTGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCTCAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATTCACCGACGCTTTCGAAGTTATG
GCTGAAAACTACGAATTCAAAGAAAACGAAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTGCCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGGTGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAA
TCTTCTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAGTTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCGTATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAACTGTCTAAAATGCAGAAAGCT
GGTCTGCTGTACTACGAAGACCTGGTTTCTTGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACACCGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCCGTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACC
AACCCGGGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAAAAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTAAAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAATGCTTCGCTATCCTGAAA
CTGTACCAGCCGCGTGTTGACAACTCTACCGCTAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTATCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTGCTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGT
CGTTACGCTGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTCGTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCACCTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCTtag

> ThetaDtD+STN

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtGAAAACCTGTACTTCCAG

ATGGAAGAAAAAGTTGGTAACCTGAAACCGAACATGGAATCTGTTAACGTTACCGTTCGTGTTCTGGAAGCTTCTGAAGCTCGTCAGATCCAGACCAAAAACGGTGTTCGTACCATCTCTGAAGCTATCGTTGGTGACGAAACCGGTCGTGTTAAACTGACCCTGTGGGGTAAA
CACGCTGGTTCTATCAAAGAAGGTCAGGTTGTTAAAATCGAAAACGCTTGGACCACCGCTTTCAAAGGTCAGGTTCAGCTGAACGCTGGTTCTAAAACCAAAATCGCTGAAGCTTCTGAAGACGGTTTCCCGGAATCTTCTCAGATCCCGGAAAACACCCCGACCGCTCCGCAG
CAGATGCGTGGTGGTGGTCGTGGTTTCCGTGGTGGTGGTCGTCGTTACGGTCGTCGTGGTGGTCGTCGTCAGGAAAACGAAGAAGGTGAAGAAGAA

ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAAAGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTAACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATGTTCCTGATGGAACTGGGTCGTCGT
AAAGGTTTCCGTGTTGAATCTGAACTGTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAGTTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAACTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCG
GTTGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGCTGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCTCAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATTCACCGACGCTTTCGAAGTTATG
GCTGAAAACTACGAATTCAAAGAAAACGAAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTGCCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGGTGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAA
TCTTCTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAGTTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCGTATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAACTGTCTAAAATGCAGAAAGCT
GGTCTGCTGTACTACGAAGACCTGGTTTCTTGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACACCGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCCGTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACC
AACCCGGGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAAAAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTAAAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAATGCTTCGCTATCCTGAAA
CTGTACCAGCCGCGTGTTGACAACTCTACCGCTAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTATCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTGCTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGT
CGTTACGCTGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTCGTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCACCTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCT

> ThetaDtD+STC

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtGAAAACCTGTACTTCCAG

ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAAAGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTAACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATGTTCCTGATGGAACTGGGTCGTCGT
AAAGGTTTCCGTGTTGAATCTGAACTGTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAGTTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAACTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCG
GTTGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGCTGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCTCAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATTCACCGACGCTTTCGAAGTTATG
GCTGAAAACTACGAATTCAAAGAAAACGAAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTGCCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGGTGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAA
TCTTCTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAGTTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCGTATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAACTGTCTAAAATGCAGAAAGCT
GGTCTGCTGTACTACGAAGACCTGGTTTCTTGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACACCGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCCGTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACC
AACCCGGGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAAAAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTAAAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAATGCTTCGCTATCCTGAAA
CTGTACCAGCCGCGTGTTGACAACTCTACCGCTAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTATCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTGCTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGT
CGTTACGCTGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTCGTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCACCTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCT

ATGGAAGAAAAAGTTGGTAACCTGAAACCGAACATGGAATCTGTTAACGTTACCGTTCGTGTTCTGGAAGCTTCTGAAGCTCGTCAGATCCAGACCAAAAACGGTGTTCGTACCATCTCTGAAGCTATCGTTGGTGACGAAACCGGTCGTGTTAAACTGACCCTGTGGGGTAAA
CACGCTGGTTCTATCAAAGAAGGTCAGGTTGTTAAAATCGAAAACGCTTGGACCACCGCTTTCAAAGGTCAGGTTCAGCTGAACGCTGGTTCTAAAACCAAAATCGCTGAAGCTTCTGAAGACGGTTTCCCGGAATCTTCTCAGATCCCGGAAAACACCCCGACCGCTCCGCAG
CAGATGCGTGGTGGTGGTCGTGGTTTCCGTGGTGGTGGTCGTCGTTACGGTCGTCGTGGTGGTCGTCGTCAGGAAAACGAAGAAGGTGAAGAAGAAtag

> ThetaDtD+SN

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtGAAAACCTGTACTTCCAG

GCTACCGTTAAATTCAAATACAAAGGTGAAGAAAAAGAAGTTGACATCTCTAAAATCAAAAAAGTTTGGCGTGTTGGTAAAATGATCTCTTTCACCTACGACGAAGGTGGTGGTAAAACCGGTCGTGGTGCTGTTTCTGAAAAAGACGCTCCGAAAGAACTGCTGCAGATGCTG
GAAAAACAGAAAAAA

ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAAAGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTAACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATGTTCCTGATGGAACTGGGTCGTCGT
AAAGGTTTCCGTGTTGAATCTGAACTGTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAGTTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAACTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCG
GTTGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGCTGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCTCAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATTCACCGACGCTTTCGAAGTTATG
GCTGAAAACTACGAATTCAAAGAAAACGAAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTGCCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGGTGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAA
TCTTCTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAGTTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCGTATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAACTGTCTAAAATGCAGAAAGCT
GGTCTGCTGTACTACGAAGACCTGGTTTCTTGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACACCGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCCGTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACC
AACCCGGGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAAAAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTAAAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAATGCTTCGCTATCCTGAAA
CTGTACCAGCCGCGTGTTGACAACTCTACCGCTAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTATCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTGCTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGT
CGTTACGCTGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTCGTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCACCTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCTtag

> ThetaDtD+SC

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtGAAAACCTGTACTTCCAG

ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAAAGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTAACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATGTTCCTGATGGAACTGGGTCGTCGT
AAAGGTTTCCGTGTTGAATCTGAACTGTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAGTTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAACTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCG
GTTGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGCTGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCTCAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATTCACCGACGCTTTCGAAGTTATG
GCTGAAAACTACGAATTCAAAGAAAACGAAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTGCCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGGTGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAA
TCTTCTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAGTTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCGTATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAACTGTCTAAAATGCAGAAAGCT
GGTCTGCTGTACTACGAAGACCTGGTTTCTTGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACACCGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCCGTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACC
AACCCGGGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAAAAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTAAAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAATGCTTCGCTATCCTGAAA
CTGTACCAGCCGCGTGTTGACAACTCTACCGCTAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTATCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTGCTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGT
CGTTACGCTGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTCGTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCACCTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCT

GCTACCGTTAAATTCAAATACAAAGGTGAAGAAAAAGAAGTTGACATCTCTAAAATCAAAAAAGTTTGGCGTGTTGGTAAAATGATCTCTTTCACCTACGACGAAGGTGGTGGTAAAACCGGTCGTGGTGCTGTTTCTGAAAAAGACGCTCCGAAAGAACTGCTGCAGATGCTG
GAAAAACAGAAAAAAtag

Codon optimized protein sequences

```
> Pfu
MDRFKAPAVISQRKRQKGLHSPKLSASYEIKFSNFVIFIMQRKMGLTRRM     50
FLMELGRRKGFRVESELSDSVTHIVAENNSYLEVLDWLKGQAVGDSSRFE     100
LLDISWFTAAMEAGRPVDSEVKYRLMEQSQSLPLNMPALEMPAFIATKVS     150
QYSCQRKTTLNNYNKKFTDAFEVMAENYEFKENEIFALEFLRAASLLKSL     200
PFSVTRMKDIQGLPAVGDQVRDIIEEIIEEGESSRVNEVLNDERYKAFKQ     250
FTSVFGVGVKTSEKWYRMGLRTVEEVKADKTLKLSKMQKAGLLYYEDLVS     300
CVSKAEADAVSLIVKNTVATFLPDALVTITGGFRLGGNIGHDIDFLITNP     350
GPREDDELLHKVIDLWKKQGLLLYADIIESTFVKEQLPSRKVDAMDHFQK     400
CFAILKLYQPRVDNSTANTSEQLEMAEVKDWKAIRVDLVITPFEQYPYAL     450
```

```
LGWTGSRQFGRDLRRYAAHERKMILDNHGLYDRRKRIFLKAGSEEEIFAH     500
LGLDYVEPWERNA

Improved DNA:
ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAA     50
AGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTA     100
ACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATG     150
TTCCTGATGGAACTGGGTCGTCGTAAAGGTTTCCGTGTTGAATCTGAACT     200
GTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAG     250
TTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAA     300
CTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCGGT     350
TGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGC     400
TGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCT     450
CAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATT     500
CACCGACGCTTTCGAAGTTATGGCTGAAAACTACGAATTCAAAGAAAACG     550
AAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTG     600
CCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGG     650
TGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAATCTT     700
CTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAG     750
TTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCG     800
TATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAAC     850
TGTCTAAAATGCAGAAAGCTGGTCTGCTGTACTACGAAGACCTGGTTTCT     900
TGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACAC     950
CGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCC     1000
GTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACCAACCCG     1050
GGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAA     1100
AAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTA     1150
AAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAA     1200
TGCTTCGCTATCCTGAAACTGTACCAGCCGCGTGTTGACAACTCTACCGC     1250
TAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTA     1300
TCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTG     1350
CTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGTCGTTACGC     1400
TGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTC     1450
GTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCAC     1500
CTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCT

> Taq (delta289)
SPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPY     50
KALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEG     100
VARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSA     150
VLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQ     200
LERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELT     250
KLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTP     300
LGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIH     350
TETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEA     400
QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKS     450
VREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVL     500
EAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE
Improved DNA:
TCTCCGAAAGCTCTGGAAGAAGCTCCGTGGCCGCCGCCGGAAGGTGCTTT     50
CGTTGGTTTCGTTCTGTCTCGTAAAGAACCGATGTGGGCTGACCTGCTGG     100
```

```
CTCTGGCTGCTGCTCGTGGTGGTCGTGTTCACCGTGCTCCGGAACCGTAC     150
AAAGCTCTGCGTGACCTGAAAGAAGCTCGTGGTCTGCTGGCTAAAGACCT     200
GTCTGTTCTGGCTCTGCGTGAAGGTCTGGGTCTGCCGCCGGGTGACGACC     250
CGATGCTGCTGGCTTACCTGCTGGACCCGTCTAACACCACCCCGGAAGGT     300
GTTGCTCGTCGTTACGGTGGTGAATGGACCGAAGAAGCTGGTGAACGTGC     350
TGCTCTGTCTGAACGTCTGTTCGCTAACCTGTGGGGTCGTCTGGAAGGTG     400
AAGAACGTCTGCTGTGGCTGTACCGTGAAGTTGAACGTCCGCTGTCTGCT     450
GTTCTGGCTCACATGGAAGCTACCGGTGTTCGTCTGGACGTTGCTTACCT     500
GCGTGCTCTGTCTCTGGAAGTTGCTGAAGAAATCGCTCGTCTGGAAGCTG     550
AAGTTTTCCGTCTGGCTGGTCACCCGTTCAACCTGAACTCTCGTGACCAG     600 CTGGAACGTGTTCTGTTCGACGAACTGGGTCTGCCGGCTATCGGTAAAAC     650 CGAAAAAACCGGTAAACGTTCTACCTCTGCTGCTGTTCTGGAAGCTCTGC     700
GTGAAGCTCACCCGATCGTTGAAAAAATCCTGCAGTACCGTGAACTGACC     750 AAACTGAAATCTACCTACATCGACCCGCTGCCGGACCTGATCCACCCGCG     800 TACCGGTCGTCTGCACACCCGTTTCAACCAGACCGCTACCGCTACCGGTC     850
GTCTGTCTTCTTCTGACCCGAACCTGCAGAACATCCCGGTTCGTACCCCG     900
CTGGGTCAGCGTATCCGTCGTGCTTTCATCGCTGAAGAAGGTTGGCTGCT     950
GGTTGCTCTGGACTACTCTCAGATCGAACTGCGTGTTCTGGCTCACCTGT     1000
CTGGTGACGAAAACCTGATCCGTGTTTTCCAGGAAGGTCGTGACATCCAC     1050
ACCGAAACCGCTTCTTGGATGTTCGGTGTTCCGCGTGAAGCTGTTGACCC     1100
GCTGATGCGTCGTGCTGCTAAAACCATCAACTTCGGTGTTCTGTACGGTA     1150
TGTCTGCTCACCGTCTGTCTCAGGAACTGGCTATCCCGTACGAAGAAGCT     1200
CAGGCTTTCATCGAACGTTACTTCCAGTCTTTCCCGAAAGTTCGTGCTTG     1250
GATCGAAAAAACCCTGGAAGAAGGTCGTCGTCGTGGTTACGTTGAAACCC     1300
TGTTCGGTCGTCGTCGTTACGTTCCGGACCTGGAAGCTCGTGTTAAATCT     1350
GTTCGTGAAGCTGCTGAACGTATGGCTTTCAACATGCCGGTTCAGGGTAC     1400
CGCTGCTGACCTGATGAAACTGGCTATGGTTAAACTGTTCCCGCGTCTGG     1450
AAGAAATGGGTGCTCGTATGCTGCTGCAGGTTCACGACGAACTGGTTCTG     1500
GAAGCTCCGAAAGAACGTGCTGAAGCTGTTGCTCGTCTGGCTAAAGAAGT     1550
TATGGAAGGTGTTTACCCGCTGGCTGTTCCGCTGGAAGTTGAAGTTGGTA     1600
TCGGTGAAGACTGGCTGTCTGCTAAAGAA
```

```
> Sso7d
ATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDA     50
PKELLQMLEKQKK

Improved DNA:
GCTACCGTTAAATTCAAATACAAAGGTGAAGAAAAAGAAGTTGACATCTC     50
TAAAATCAAAAAAGTTTGGCGTGTTGGTAAAATGATCTCTTTCACCTACG     100
ACGAAGGTGGTGGTAAAACCGGTCGTGGTGCTGTTTCTGAAAAAGACGCT     150
CCGAAAGAACTGCTGCAGATGCTGGAAAAACAGAAAAAA

> ThetaSsoT
MEEKVGNLKPNMESVNVTVRVLEASEARQIQTKNGVRTISEAIVGDETGR     50
VKLTLWGKHAGSIKEGQVVKIENAWTTAFKGQVQLNAGSKTKIAEASEDG     100
FPESSQIPENTPTAPQQMRGGGRGFRGGGRRYGRRGGRRQENEEGEEE

Improved DNA:
ATGGAAGAAAAAGTTGGTAACCTGAAACCGAACATGGAATCTGTTAACGT     50
TACCGTTCGTGTTCTGGAAGCTTCTGAAGCTCGTCAGATCCAGACCAAAA     100
ACGGTGTTCGTACCATCTCTGAAGCTATCGTTGGTGACGAAACCGGTCGT     150
GTTAAACTGACCCTGTGGGGTAAACACGCTGGTTCTATCAAAGAAGGTCA     200
GGTTGTTAAAATCGAAAACGCTTGGACCACCGCTTTCAAAGGTCAGGTTC     250
AGCTGAACGCTGGTTCTAAAACCAAAATCGCTGAAGCTTCTGAAGACGGT     300
```

TTCCCGGAATCTTCTCAGATCCCGGAAAACACCCCGACCGCTCCGCAGCA 350
GATGCGTGGTGGTGGTCGTGGTTTCCGTGGTGGTGGTCGTCGTTACGGTC 400
GTCGTGGTGGTCGTCGTCAGGAAAACGAAGAAGGTGAAGAAGAA

> Theta43 Expression construct
TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

2) ThetaPfu+S

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtGAAAACCTGTACTTCCAG

ATGGACCGTTTCAAAGCTCCGGCTGTTATCTCTCAGCGTAAACGTCAGAAAGGTCTGCACTCTCCGAAACTGTCTGCTTCTTACGAAATCAAATTCTCTAACTTCGTTATCTTCATCATGCAGCGTAAAATGGGTCTGACCCGTCGTATGTTCCTGATGGAACTGGGTCGTCGT
AAAGGTTTCCGTGTTGAATCTGAACTGTCTGACTCTGTTACCCACATCGTTGCTGAAAACAACTCTTACCTGGAAGTTCTGGACTGGCTGAAAGGTCAGGCTGTTGGTGACTCTTCTCGTTTCGAACTGCTGGACATCTCTTGGTTCACCGCTGCTATGGAAGCTGGTCGTCCG
GTTGACTCTGAAGTTAAATACCGTCTGATGGAACAGTCTCAGTCTCTGCCGCTGAACATGCCGGCTCTGGAAATGCCGGCTTTCATCGCTACCAAAGTTTCTCAGTACTCTTGCCAGCGTAAAACCACCCTGAACAACTACAACAAAAAATTCACCGACGCTTTCGAAGTTATG
GCTGAAAACTACGAATTCAAAGAAAACGAAATCTTCGCTCTGGAATTCCTGCGTGCTGCTTCTCTGCTGAAATCTCTGCCGTTCTCTGTTACCCGTATGAAAGACATCCAGGGTCTGCCGGCTGTTGGTGACCAGGTTCGTGACATCATCGAAGAAATCATCGAAGAAGGTGAA
TCTTCTCGTGTTAACGAAGTTCTGAACGACGAACGTTACAAAGCTTTCAAACAGTTCACCTCTGTTTTCGGTGTTGGTGTTAAAACCTCTGAAAAATGGTACCGTATGGGTCTGCGTACCGTTGAAGAAGTTAAAGCTGACAAAACCCTGAAACTGTCTAAAATGCAGAAAGCT
GGTCTGCTGTACTACGAAGACCTGGTTTCTTGCGTTTCTAAAGCTGAAGCTGACGCTGTTTCTCTGATCGTTAAAAACACCGTTGCTACCTTCCTGCCGGACGCTCTGGTTACCATCACCGGTGGTTTCCGTCTGGGTGGTAACATCGGTCACGACATCGACTTCCTGATCACC
AACCCGGGTCCGCGTGAAGACGACGAACTGCTGCACAAAGTTATCGACCTGTGGAAAAAACAGGGTCTGCTGCTGTACGCTGACATCATCGAATCTACCTTCGTTAAAGAACAGCTGCCGTCTCGTAAAGTTGACGCTATGGACCACTTCCAGAAATGCTTCGCTATCCTGAAA
CTGTACCAGCCGCGTGTTGACAACTCTACCGCTAACACCTCTGAACAGCTGGAAATGGCTGAAGTTAAAGACTGGAAAGCTATCCGTGTTGACCTGGTTATCACCCCGTTCGAACAGTACCCGTACGCTCTGCTGGGTTGGACCGGTTCTCGTCAGTTCGGTCGTGACCTGCGT
CGTTACGCTGCTCACGAACGTAAAATGATCCTGGACAACCACGGTCTGTACGACCGTCGTAAACGTATCTTCCTGAAAGCTGGTTCTGAAGAAGAAATCTTCGCTCACCTGGGTCTGGACTACGTTGAACCGTGGGAACGTAACGCT

GCTACCGTTAAATTCAAATACAAAGGTGAAGAAAAAGAAGTTGACATCTCTAAAATCAAAAAAGTTTGGCGTGTTGGTAAAATGATCTCTTTCACCTACGACGAAGGTGGTGGTAAAACCGGTCGTGGTGCTGTTTCTGAAAAAGACGCTCCGAAAGAACTGCTGCAGATGCTG
GAAAAACAGAAAAAAtag

TCTCCGAAAGCTCTGGAAGAAGCTCCGTGGCCGCCGCCGGAAGGTGCTTTCGTTGGTTTCGTTCTGTCTCGTAAAGAACCGATGTGGGCTGACCTGCTGGCTCTGGCTGCTGCTCGTGGTGGTCGTGTTCACCGTGCTCCGGAACCGTACAAAGCTCTGCGTGACCTGAAAGAA
GCTCGTGGTCTGCTGGCTAAAGACCTGTCTGTTCTGGCTCTGCGTGAAGGTCTGGGTCTGCCGCCGGGTGACGACCCGATGCTGCTGGCTTACCTGCTGGACCCGTCTAACACCACCCCGGAAGGTGTTGCTCGTCGTTACGGTGGTGAATGGACCGAAGAAGCTGGTGAACGT
GCTGCTCTGTCTGAACGTCTGTTCGCTAACCTGTGGGGTCGTCTGGAAGGTGAAGAACGTCTGCTGTGGCTGTACCGTGAAGTTGAACGTCCGCTGTCTGCTGTTCTGGCTCACATGGAAGCTACCGGTGTTCGTCTGGACGTTGCTTACCTGCGTGCTCTGTCTCTGGAAGTT
GCTGAAGAAATCGCTCGTCTGGAAGCTGAAGTTTTCCGTCTGGCTGGTCACCCGTTCAACCTGAACTCTCGTGACCAGCTGGAACGTGTTCTGTTCGACGAACTGGGTCTGCCGGCTATCGGTAAAACCGAAAAAACCGGTAAACGTTCTACCTCTGCTGCTGTTCTGGAAGCT
CTGCGTGAAGCTCACCCGATCGTTGAAAAAATCCTGCAGTACCGTGAACTGACCAAACTGAAATCTACCTACATCGACCCGCTGCCGGACCTGATCCACCCGCGTACCGGTCGTCTGCACACCCGTTTCAACCAGACCGCTACCGCTACCGGTCGTCTGTCTTCTTCTGACCCG
AACCTGCAGAACATCCCGGTTCGTACCCCGCTGGGTCAGCGTATCCGTCGTGCTTTCATCGCTGAAGAAGGTTGGCTGCTGGTTGCTCTGGACTACTCTCAGATCGAACTGCGTGTTCTGGCTCACCTGTCTGGTGACGAAAACCTGATCCGTGTTTTCCAGGAAGGTCGTGAC
ATCCACACCGAAACCGCTTCTTGGATGTTCGGTGTTCCGCGTGAAGCTGTTGACCCGCTGATGCGTCGTGCTGCTAAAACCATCAACTTCGGTGTTCTGTACGGTATGTCTGCTCACCGTCTGTCTCAGGAACTGGCTATCCCGTACGAAGAAGCTCAGGCTTTCATCGAACGT
TACTTCCAGTCTTTCCCGAAAGTTCGTGCTTGGATCGAAAAAACCCTGGAAGAAGGTCGTCGTCGTGGTTACGTTGAAACCCTGTTCGGTCGTCGTCGTTACGTTCCGGACCTGGAAGCTCGTGTTAAATCTGTTCGTGAAGCTGCTGAACGTATGGCTTTCAACATGCCGGTT
CAGGGTACCGCTGCTGACCTGATGAAACTGGCTATGGTTAAACTGTTCCCGCGTCTGGAAGAAATGGGTGCTCGTATGCTGCTGCAGGTTCACGACGAACTGGTTCTGGAAGCTCCGAAAGAACGTGCTGAAGCTGTTGCTCGTCTGGCTAAAGAAGTTATGGAAGGTGTTTAC
CCGCTGGCTGTTCCGCTGGAAGTTGAAGTTGGTATCGGTGAAGACTGGCTGTCTGCTAAAGAAtag

4) Theta+PfUTPase

```
> PfUTPase
MLHHVKLIYATKSRKLVGKKIVLAIPGSIAAVECVKLARELIRHGAEVHA     50
VMSEAATKIIHPYAMEFATGNPVITEITGFIEHVELAGEHENKADLILVC     100
PATANTISKIACGIDDTPVTTVVTTAFPHIPIMIAPAMHETMYRHPIVRE     150
NIERLKKLGVEFIGPRIEEGKAKVASIDEIVYRVIKKLHKKTLEGKRVLV     200
TAGATREYIDPIRFITNASSGKMGVALAEEADFRGAEVTLIRTKGSVKSF     250
VENQIEVETVEEMLSAIENELRSKKYDVVIMAAAVSDFRPKIKAEGKIKS     300
DRSITIELVPNPKIIDRIKEIQPNVFLVGFKAETSKEKLIEEGKRQIERA     350
KADLVVGNTLEAFGSEENQVVLIGRDFTKELPKMKKRELAERIWDEIEKL     400
LS

Improved DNA:
ATGCTGCACCACGTTAAACTGATCTACGCTACCAAATCTCGTAAACTGGT     50
TGGTAAAAAAATCGTTCTGGCTATCCCGGGTTCTATCGCTGCTGTTGAAT     100
GCGTTAAACTGGCTCGTGAACTGATCCGTCACGGTGCTGAAGTTCACGCT     150
GTTATGTCTGAAGCTGCTACCAAAATCATCCACCCGTACGCTATGGAATT     200
CGCTACCGGTAACCCGGTTATCACCGAAATCACCGGTTTCATCGAACACG     250
TTGAACTGGCTGGTGAACACGAAAACAAAGCTGACCTGATCCTGGTTTGC     300
CCGGCTACCGCTAACACCATCTCTAAAATCGCTTGCGGTATCGACGACAC     350
CCCGGTTACCACCGTTGTTACCACCGCTTTCCCGCACATCCCGATCATGA     400
TCGCTCCGGCTATGCACGAAACCATGTACCGTCACCCGATCGTTCGTGAA     450
AACATCGAACGTCTGAAAAAACTGGGTGTTGAATTCATCGGTCCGCGTAT     500
CGAAGAAGGTAAAGCTAAAGTTGCTTCTATCGACGAAATCGTTTACCGTG     550
TTATCAAAAAACTGCACAAAAAAACCCTGGAAGGTAAACGTGTTCTGGTT     600
ACCGCTGGTGCTACCCGTGAATACATCGACCCGATCCGTTTCATCACCAA     650
CGCTTCTTCTGGTAAAATGGGTGTTGCTCTGGCTGAAGAAGCTGACTTCC     700
GTGGTGCTGAAGTTACCCTGATCCGTACCAAAGGTTCTGTTAAATCTTTC     750
GTTGAAAACCAGATCGAAGTTGAAACCGTTGAAGAAATGCTGTCTGCTAT     800
CGAAAACGAACTGCGTTCTAAAAAATACGACGTTGTTATCATGGCTGCTG     850
CTGTTTCTGACTTCCGTCCGAAAATCAAAGCTGAAGGTAAAATCAAATCT     900
GACCGTTCTATCACCATCGAACTGGTTCCGAACCCGAAAATCATCGACCG     950
TATCAAAGAAATCCAGCCGAACGTTTTCCTGGTTGGTTTCAAAGCTGAAA     1000
CCTCTAAAGAAAAACTGATCGAAGAAGGTAAACGTCAGATCGAACGTGCT     1050
AAAGCTGACCTGGTTGTTGGTAACACCCTGGAAGCTTTCGGTTCTGAAGA     1100
AAACCAGGTTGTTCTGATCGGTCGTGACTTCACCAAAGAACTGCCGAAAA     1150
TGAAAAAACGTGAACTGGCTGAACGTATCTGGGACGAAATCGAAAAACTG     1200
CTGTCT

> Theta+PfUTPase
TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

ATGCTGCACCACGTTAAACTGATCTACGCTACCAAATCTCGTAAACTGGTTGGTAAAAAAATCGTTCTGGCTATCCCGGGTTCTATCGCTGCTGTTGAATGCGTTAAACTGGCTCGTGAACTGATCCGTCACGGTGCTGAAGTTCACGCTGTTATGTCTGAAGCTGCTACCAAA
ATCATCCACCCGTACGCTATGGAATTCGCTACCGGTAACCCGGTTATCACCGAAATCACCGGTTTCATCGAACACGTTGAACTGGCTGGTGAACACGAAAACAAAGCTGACCTGATCCTGGTTTGCCCGGCTACCGCTAACACCATCTCTAAAATCGCTTGCGGTATCGACGAC
ACCCCGGTTACCACCGTTGTTACCACCGCTTTCCCGCACATCCCGATCATGATCGCTCCGGCTATGCACGAAACCATGTACCGTCACCCGATCGTTCGTGAAAACATCGAACGTCTGAAAAAACTGGGTGTTGAATTCATCGGTCCGCGTATCGAAGAAGGTAAAGCTAAAGTT
GCTTCTATCGACGAAATCGTTTACCGTGTTATCAAAAAACTGCACAAAAAAACCCTGGAAGGTAAACGTGTTCTGGTTACCGCTGGTGCTACCCGTGAATACATCGACCCGATCCGTTTCATCACCAACGCTTCTTCTGGTAAAATGGGTGTTGCTCTGGCTGAAGAAGCTGAC
TTCCGTGGTGCTGAAGTTACCCTGATCCGTACCAAAGGTTCTGTTAAATCTTTCGTTGAAAACCAGATCGAAGTTGAAACCGTTGAAGAAATGCTGTCTGCTATCGAAAACGAACTGCGTTCTAAAAAATACGACGTTGTTATCATGGCTGCTGCTGTTTCTGACTTCCGTCCG
AAAATCAAAGCTGAAGGTAAAATCAAATCTGACCGTTCTATCACCATCGAACTGGTTCCGAACCCGAAAATCATCGACCGTATCAAAGAAATCCAGCCGAACGTTTTCCTGGTTGGTTTCAAAGCTGAAACCTCTAAAGAAAAACTGATCGAAGAAGGTAAACGTCAGATCGAA
```

CGTGCTAAAGCTGACCTGGTTGTTGGTAACACCCTGGAAGCTTTCGGTTCTGAAGAAAACCAGGTTGTTCTGATCGGTCGTGACTTCACCAAAGAACTGCCGAAAATGAAAAAACGTGAACTGGCTGAACGTATCTGGGACGAAATCGAAAAACTGCTGTCTtag

5) Theta+TEV

> TEV protease
Translation:
MSLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRN 50
NGTLLVQSLHGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFR 100
EPQREERICLVTTNFQTKSMSSMVSDTSCTFPSSDGIFWKHWIQTKDGQC 150
GSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSG 200
WRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMNELVYSQ

Improved DNA:
ATGTCTCTGTTCAAAGGTCCGCGTGACTACAACCCGATCTCTTCTACCAT 50 CTGCCACCTGACCAACGAATCTGACGGTCACACCACCTCTCTGTACGGTA
100 TCGGTTTCGGTCCGTTCATCATCACCAACAAACACCTGTTCCGTCGTAAC 150
AACGGTACCCTGCTGGTTCAGTCTCTGCACGGTGTTTTCAAAGTTAAAAA 200
CACCACCACCCTGCAGCAGCACCTGATCGACGGTCGTGACATGATCATCA 250
TCCGTATGCCGAAAGACTTCCCGCCGTTCCCGCAGAAACTGAAATTCCGT 300
GAACCGCAGCGTGAAGAACGTATCTGCCTGGTTACCACCAACTTCCAGAC 350
CAAATCTATGTCTTCTATGGTTTCTGACACCTCTTGCACCTTCCCGTCTT 400
CTGACGGTATCTTCTGGAAACACTGGATCCAGACCAAAGACGGTCAGTGC 450
GGTTCTCCGCTGGTTTCTACCCGTGACGGTTTCATCGTTGGTATCCACTC 500
TGCTTCTAACTTCACCAACACCAACAACTACTTCACCTCTGTTCCGAAAA 550
ACTTCATGGAACTGCTGACCAACCAGGAAGCTCAGCAGTGGGTTTCTGGT 600
TGGCGTCTGAACGCTGACTCTGTTCTGTGGGGTGGTCACAAAGTTTTCAT 650
GGTTAAACCGGAAGAACCGTTCCAGCCGGTTAAAGAAGCTACCCAGCTGA 700
TGAACGAACTGGTTTACTCTCAG

> ThetaTEV
TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

ATGTCTCTGTTCAAAGGTCCGCGTGACTACAACCCGATCTCTTCTACCATCTGCCACCTGACCAACGAATCTGACGGTCACACCACCTCTCTGTACGGTATCGGTTTCGGTCCGTTCATCATCACCAACAAACACCTGTTCCGTCGTAACAACGGTACCCTGCTGGTTCAGTCT
CTGCACGGTGTTTTCAAAGTTAAAAACACCACCACCCTGCAGCAGCACCTGATCGACGGTCGTGACATGATCATCATCCGTATGCCGAAAGACTTCCCGCCGTTCCCGCAGAAACTGAAATTCCGTGAACCGCAGCGTGAAGAACGTATCTGCCTGGTTACCACCAACTTCCAG
ACCAAATCTATGTCTTCTATGGTTTCTGACACCTCTTGCACCTTCCCGTCTTCTGACGGTATCTTCTGGAAACACTGGATCCAGACCAAAGACGGTCAGTGCGGTTCTCCGCTGGTTTCTACCCGTGACGGTTTCATCGTTGGTATCCACTCTGCTTCTAACTTCACCAACACC
AACAACTACTTCACCTCTGTTCCGAAAAACTTCATGGAACTGCTGACCAACCAGGAAGCTCAGCAGTGGGTTTCTGGTTGGCGTCTGAACGCTGACTCTGTTCTGTGGGGTGGTCACAAAGTTTTCATGGTTAAACCGGAAGAACCGTTCCAGCCGGTTAAAGAAGCTACCCAG
CTGATGAACGAACTGGTTTACTCTCAGtag

6) Theta+SsoT

> Theta+SsoT
TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

ATGGAAGAAAAAGTTGGTAACCTGAAACCGAACATGGAATCTGTTAACGTTACCGTTCGTGTTCTGGAAGCTTCTGAAGCTCGTCAGATCCAGACCAAAAACGGTGTTCGTACCATCTCTGAAGCTATCGTTGGTGACGAAACCGGTCGTGTTAAACTGACCCTGTGGGGTAAA
CACGCTGGTTCTATCAAAGAAGGTCAGGTTGTTAAAATCGAAAACGCTTGGACCACCGCTTTCAAAGGTCAGGTTCAGCTGAACGCTGGTTCTAAAACCAAAATCGCTGAAGCTTCTGAAGACGGTTTCCCGGAATCTTCTCAGATCCCGGAAAACACCCCGACCGCTCCGCAG
CAGATGCGTGGTGGTGGTCGTGGTTTCCGTGGTGGTGGTCGTCGTTACGGTCGTCGTGGTGGTCGTCGTCAGGAAAACGAAGAAGGTGAAGAAGAAtag

7) Theta+ISO

> Theta+ISO

```
MEAKGEKPLEEMEFAIVDVITEEMLADKAALVVEVMEENYHDAPIVGIAL     50
VNEHGRFFMRPETALADSQFLAWLADETKKKSMFDAKRAVVALKWKGIEL     100
RGVAFDLLLAAYLLNPAQDAGDIAAVAKMKQYEAVRSDEAVYGKGVKRSL     150
PDEQTLAEHLVRKAAAIWALEQPFMDDLRNNEQDQLLTKLEHALAAILAE     200
MEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYELAGQEFNINSPKQLGVI     250
LFEKLQLPVLKKTKTGYSTSADVLEKLAPHHEIVENILHYRQLGKLQSTY     300
IEGLLKVVRPDTGKVHTMFNQALTQTGRLSSAEPNLQNIPIRLEEGRKIR     350
QAFVPSEPDWLIFAADYSQIELRVLAHIADDDNLIEAFQRDLDIHTKTAM     400
DIFQLSEEEVTANMRRQAKAVNFGIVYGISDYGLAQNLNITRKEAAEFIE     450
RYFASFPGVKQYMENIVQEAKQKGYVTTLLHRRRYLPDITSRNFNVRSFA     500
ERTAMNTPIQGSAADIIKKAMIDLAARLKEEQLQARLLLQVHDELILEAP     550
KEEIERLCELVPEVMEQAVTLRVPLKVDYHYGPTWYDAK
```

Improved DNA:

```
ATGGAAGCTAAAGGTGAAAAACCGCTGGAAGAAATGGAATTCGCTATCGT     50
TGACGTTATCACCGAAGAAATGCTGGCTGACAAAGCTGCTCTGGTTGTTG     100
AAGTTATGGAAGAAAACTACCACGACGCTCCGATCGTTGGTATCGCTCTG     150
GTTAACGAACACGGTCGTTTCTTCATGCGTCCGGAAACCGCTCTGGCTGA     200
CTCTCAGTTCCTGGCTTGGCTGGCTGACGAAACCAAAAAAAAATCTATGT     250
TCGACGCTAAACGTGCTGTTGTTGCTCTGAAATGGAAAGGTATCGAACTG     300
CGTGGTGTTGCTTTCGACCTGCTGCTGGCTGCTTACCTGCTGAACCCGGC     350
TCAGGACGCTGGTGACATCGCTGCTGTTGCTAAAATGAAACAGTACGAAG     400
CTGTTCGTTCTGACGAAGCTGTTTACGGTAAAGGTGTTAAACGTTCTCTG     450
CCGGACGAACAGACCCTGGCTGAACACCTGGTTCGTAAAGCTGCTGCTAT     500
CTGGGCTCTGGAACAGCCGTTCATGGACGACCTGCGTAACAACGAACAGG     550
ACCAGCTGCTGACCAAACTGGAACACGCTCTGGCTGCTATCCTGGCTGAA     600
ATGGAATTCACCGGTGTTAACGTTGACACCAAACGTCTGGAACAGATGGG     650
TTCTGAACTGGCTGAACAGCTGCGTGCTATCGAACAGCGTATCTACGAAC     700
TGGCTGGTCAGGAATTCAACATCAACTCTCCGAAACAGCTGGGTGTTATC     750
CTGTTCGAAAAACTGCAGCTGCCGGTTCTGAAAAAAACCAAAACCGGTTA     800
CTCTACCTCTGCTGACGTTCTGGAAAAACTGGCTCCGCACCACGAAATCG     850
TTGAAAACATCCTGCACTACCGTCAGCTGGGTAAACTGCAGTCTACCTAC     900
ATCGAAGGTCTGCTGAAAGTTGTTCGTCCGGACACCGGTAAAGTTCACAC     950
CATGTTCAACCAGGCTCTGACCCAGACCGGTCGTCTGTCTTCTGCTGAAC     1000
CGAACCTGCAGAACATCCCGATCCGTCTGGAAGAAGGTCGTAAAATCCGT     1050
CAGGCTTTCGTTCCGTCTGAACCGGACTGGCTGATCTTCGCTGCTGACTA     1100
CTCTCAGATCGAACTGCGTGTTCTGGCTCACATCGCTGACGACGACAACC     1150
TGATCGAAGCTTTCCAGCGTGACCTGGACATCCACACCAAAACCGCTATG     1200
GACATCTTCCAGCTGTCTGAAGAAGAAGTTACCGCTAACATGCGTCGTCA     1250
GGCTAAAGCTGTTAACTTCGGTATCGTTTACGGTATCTCTGACTACGGTC     1300
TGGCTCAGAACCTGAACATCACCCGTAAAGAAGCTGCTGAATTCATCGAA     1350
CGTTACTTCGCTTCTTTCCCGGGTGTTAAACAGTACATGGAAAACATCGT     1400
TCAGGAAGCTAAACAGAAAGGTTACGTTACCACCCTGCTGCACCGTCGTC     1450
GTTACCTGCCGGACATCACCTCTCGTAACTTCAACGTTCGTTCTTTCGCT     1500
GAACGTACCGCTATGAACACCCCGATCCAGGGTTCTGCTGCTGACATCAT     1550
CAAAAAAGCTATGATCGACCTGGCTGCTCGTCTGAAAGAAGAACAGCTGC     1600
AGGCTCGTCTGCTGCTGCAGGTTCACGACGAACTGATCCTGGAAGCTCCG     1650
AAAGAAGAAATCGAACGTCTGTGCGAACTGGTTCCGGAAGTTATGGAACA     1700
```

GGCTGTTACCCTGCGTGTTCCGCTGAAAGTTGACTACCACTACGGTCCGA 1750
CCTGGTACGACGCTAAA

> Theta+ISO
TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

ATGGAAGCTAAAGGTGAAAAACCGCTGGAAGAAATGGAATTCGCTATCGTTGACGTTATCACCGAAGAAATGCTGGCTGACAAAGCTGCTCTGGTTGTTGAAGTTATGGAAGAAAACTACCACGACGCTCCGATCGTTGGTATCGCTCTGGTTAACGAACACGGTCGTTTCTTC
ATGCGTCCGGAAACCGCTCTGGCTGACTCTCAGTTCCTGGCTTGGCTGGCTGACGAAACCAAAAAAAAATCTATGTTCGACGCTAAACGTGCTGTTGTTGCTCTGAAATGGAAAGGTATCGAACTGCGTGGTGTTGCTTTCGACCTGCTGCTGGCTGCTTACCTGCTGAACCCG
GCTCAGGACGCTGGTGACATCGCTGCTGTTGCTAAAATGAAACAGTACGAAGCTGTTCGTTCTGACGAAGCTGTTTACGGTAAAGGTGTTAAACGTTCTCTGCCGGACGAACAGACCCTGGCTGAACACCTGGTTCGTAAAGCTGCTGCTATCTGGGCTCTGGAACAGCCGTTC
ATGGACGACCTGCGTAACAACGAACAGGACCAGCTGCTGACCAAACTGGAACACGCTCTGGCTGCTATCCTGGCTGAAATGGAATTCACCGGTGTTAACGTTGACACCAAACGTCTGGAACAGATGGGTTCTGAACTGGCTGAACAGCTGCGTGCTATCGAACAGCGTATCTAC
GAACTGGCTGGTCAGGAATTCAACATCAACTCTCCGAAACAGCTGGGTGTTATCCTGTTCGAAAAACTGCAGCTGCCGGTTCTGAAAAAAACCAAAACCGGTTACTCTACCTCTGCTGACGTTCTGGAAAAACTGGCTCCGCACCACGAAATCGTTGAAAACATCCTGCACTAC
CGTCAGCTGGGTAAACTGCAGTCTACCTACATCGAAGGTCTGCTGAAAGTTGTTCGTCCGGACACCGGTAAAGTTCACACCATGTTCAACCAGGCTCTGACCCAGACCGGTCGTCTGTCTTCTGCTGAACCGAACCTGCAGAACATCCCGATCCGTCTGGAAGAAGGTCGTAAA
ATCCGTCAGGCTTTCGTTCCGTCTGAACCGGACTGGCTGATCTTCGCTGCTGACTACTCTCAGATCGAACTGCGTGTTCTGGCTCACATCGCTGACGACGACAACCTGATCGAAGCTTTCCAGCGTGACCTGGACATCCACACCAAAACCGCTATGGACATCTTCCAGCTGTCT
GAAGAAGAAGTTACCGCTAACATGCGTCGTCAGGCTAAAGCTGTTAACTTCGGTATCGTTTACGGTATCTCTGACTACGGTCTGGCTCAGAACCTGAACATCACCCGTAAAGAAGCTGCTGAATTCATCGAACGTTACTTCGCTTCTTTCCCGGGTGTTAAACAGTACATGGAA
AACATCGTTCAGGAAGCTAAACAGAAAGGTTACGTTACCACCCTGCTGCACCGTCGTCGTTACCTGCCGGACATCACCTCTCGTAACTTCAACGTTCGTTCTTTCGCTGAACGTACCGCTATGAACACCCCGATCCAGGGTTCTGCTGCTGACATCATCAAAAAAGCTATGATC
GACCTGGCTGCTCGTCTGAAAGAAGAACAGCTGCAGGCTCGTCTGCTGCTGCAGGTTCACGACGAACTGATCCTGGAAGCTCCGAAAGAAGAAATCGAACGTCTGTGCGAACTGGTTCCGGAAGTTATGGAACAGGCTGTTACCCTGCGTGTTCCGCTGAAAGTTGACTACCAC
TACGGTCCGACCTGGTACGACGCTAAAtag

8) Theta+EXO

> Theta+EXO
MSKSWGKFIEEEEAEMASRRNLMIVDGTNLGFRFKHNNSKKPFASSYVST 50 IQSLAKSYSARTTIVLGDKGKSVFRLEHLPEYKGNRDEKYAQRTEEEKAL
100 DEQFFEYLKDAFELCKTTFPTFTIRGVEADDMAAYIVKLIGHLYDHVWLI 150
STDGDWDTLLTDKVSRFSFTTRREYHLRDMYEHHNVDDVEQFISLKAIMG 200
DLGDNIRGVEGIGAKRGYNIIREFGNVLDIIDQLPLPGKQKYIQNLNASE 250 ELLFRNLILVDLPTYCVDAIAAVGQDVLDKFTKDILEIAEQ

Improved DNA:
ATGTCTAAATCTTGGGGTAAATTCATCGAAGAAGAAGAAGCTGAAATGGC 50 TTCTCGTCGTAACCTGATGATCGTTGACGGTACCAACCTGGGTTTCCGTT
100 TCAAACACAACAACTCTAAAAAACCGTTCGCTTCTTCTTACGTTTCTACC 150
ATCCAGTCTCTGGCTAAATCTTACTCTGCTCGTACCACCATCGTTCTGGG 200
TGACAAAGGTAAATCTGTTTTCCGTCTGGAACACCTGCCGGAATACAAAG 250
GTAACCGTGACGAAAAATACGCTCAGCGTACCGAAGAAGAAAAAGCTCTG 300
GACGAACAGTTCTTCGAATACCTGAAAGACGCTTTCGAACTGTGCAAAAC 350
CACCTTCCCGACCTTCACCATCCGTGGTGTTGAAGCTGACGACATGGCTG 400
CTTACATCGTTAAACTGATCGGTCACCTGTACGACCACGTTTGGCTGATC 450
TCTACCGACGGTGACTGGGACACCCTGCTGACCGACAAAGTTTCTCGTTT 500
CTCTTTCACCACCCGTCGTGAATACCACCTGCGTGACATGTACGAACACC 550
ACAACGTTGACGACGTTGAACAGTTCATCTCTCTGAAAGCTATCATGGGT 600
GACCTGGGTGACAACATCCGTGGTGTTGAAGGTATCGGTGCTAAACGTGG 650
TTACAACATCATCCGTGAATTCGGTAACGTTCTGGACATCATCGACCAGC 700
TGCCGCTGCCGGGTAAACAGAAATACATCCAGAACCTGAACGCTTCTGAA 750
GAACTGCTGTTCCGTAACCTGATCCTGGTTGACCTGCCGACCTACTGCGT 800
TGACGCTATCGCTGCTGTTGGTCAGGACGTTCTGGACAAATTCACCAAAG 850
ACATCCTGGAAATCGCTGAACAG

>Theta+EXO

TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

ATGTCTAAATCTTGGGGTAAATTCATCGAAGAAGAAGAAGCTGAAATGGCTTCTCGTCGTAACCTGATGATCGTTGACGGTACCAACCTGGGTTTCCGTTTCAAACACAACAACTCTAAAAAACCGTTCGCTTCTTCTTACGTTTCTACCATCCAGTCTCTGGCTAAATCTTAC
TCTGCTCGTACCACCATCGTTCTGGGTGACAAAGGTAAATCTGTTTTCCGTCTGGAACACCTGCCGGAATACAAAGGTAACCGTGACGAAAAATACGCTCAGCGTACCGAAGAAGAAAAAGCTCTGGACGAACAGTTCTTCGAATACCTGAAAGACGCTTTCGAACTGTGCAAA
ACCACCTTCCCGACCTTCACCATCCGTGGTGTTGAAGCTGACGACATGGCTGCTTACATCGTTAAACTGATCGGTCACCTGTACGACCACGTTTGGCTGATCTCTACCGACGGTGACTGGGACACCCTGCTGACCGACAAAGTTTCTCGTTTCTCTTTCACCACCCGTCGTGAA
TACCACCTGCGTGACATGTACGAACACCACAACGTTGACGACGTTGAACAGTTCATCTCTCTGAAAGCTATCATGGGTGACCTGGGTGACAACATCCGTGGTGTTGAAGGTATCGGTGCTAAACGTGGTTACAACATCATCCGTGAATTCGGTAACGTTCTGGACATCATCGAC
CAGCTGCCGCTGCCGGGTAAACAGAAATACATCCAGAACCTGAACGCTTCTGAAGAACTGCTGTTCCGTAACCTGATCCTGGTTGACCTGCCGACCTACTGCGTTGACGCTATCGCTGCTGTTGGTCAGGACGTTCTGGACAAATTCACCAAAGACATCCTGGAAATCGCTGAA
CAGtag

9) Theta+LGT

> Theta+LGT

```
MTLEEARKRVNELRDLIRYHNYRYYVLADPEISDAEYDRLLRELKELEER     50
FPELKSPDSPTLQVGARPLEATFRPVRHPTRMYSLDNAFNLDELKAFEER     100
IERALGRKGPFAYTVEHKVDGLSVNLYYEEGVLVYGATRGDGEVGEEVTQ     150
NLLTIPTIPRRLKGVPERLEVRGEVYMPIEAFLRLNEELEERGERIFKNP     200
RNAAAGSLRQKDPRITAKRGLRATFYALGLGLEEVEREGVATQFALLHWL     250
KEKGFPVEHGYARAVGAEGVEAVYQDWLKKRRALPFEADGVVVRLDELAL     300
WRELGYTARAPRFAIAYKFPAEEKETRLLDVVFQVGRTGRVTPVGILEPV     350
FLEGSEVSRVTLHNESYIEELDIRIGDWVLVHKAGGVIPEVLRVLKERRT     400
GEERPIRWPETCPECGHRLLKEGKVHRCPNPLCPAKRFEAIRHFASRKAM     450
DIQGLGEKLIERLLEKGLVKDVADLYRLRKEDLVGLERMGEKSAQNLLRQ     500
IEESKKRGLERLLYALGLPGVGEVLARNLAARFGNMDRLLEASLEELLEV     550
EEVGELTARAILETLKDPAFRDLVRRLKEAGVEMEAKEKGGEALKGLTFV     600
ITGELSRPREEVKALLRRLGAKVTDSVSRKTSYLVVGENPGSKLEKARAL     650
GVPTLTEEELYRLLEARTGKKAEELV
```

Improved DNA:

```
ATGACCCTGGAAGAAGCTCGTAAACGTGTTAACGAACTGCGTGACCTGAT     50 CCGTTACCACAACTACCGTTACTACGTTCTGGCTGACCCGGAAATCTCTG
100
ACGCTGAATACGACCGTCTGCTGCGTGAACTGAAAGAACTGGAAGAACGT     150
TTCCCGGAACTGAAATCTCCGGACTCTCCGACCCTGCAGGTTGGTGCTCG     200
TCCGCTGGAAGCTACCTTCCGTCCGGTTCGTCACCCGACCCGTATGTACT     250
CTCTGGACAACGCTTTCAACCTGGACGAACTGAAAGCTTTCGAAGAACGT     300
ATCGAACGTGCTCTGGGTCGTAAAGGTCCGTTCGCTTACACCGTTGAACA     350
CAAAGTTGACGGTCTGTCTGTTAACCTGTACTACGAAGAAGGTGTTCTGG     400
TTTACGGTGCTACCCGTGGTGACGGTGAAGTTGGTGAAGAAGTTACCCAG     450
AACCTGCTGACCATCCCGACCATCCCGCGTCGTCTGAAAGGTGTTCCGGA     500
ACGTCTGGAAGTTCGTGGTGAAGTTTACATGCCGATCGAAGCTTTCCTGC     550
GTCTGAACGAAGAACTGGAAGAACGTGGTGAACGTATCTTCAAAAACCCG     600
CGTAACGCTGCTGCTGGTTCTCTGCGTCAGAAAGACCCGCGTATCACCGC     650
TAAACGTGGTCTGCGTGCTACCTTCTACGCTCTGGGTCTGGGTCTGGAAG     700
AAGTTGAACGTGAAGGTGTTGCTACCCAGTTCGCTCTGCTGCACTGGCTG     750
AAAGAAAAAGGTTTCCCGGTTGAACACGGTTACGCTCGTGCTGTTGGTGC     800
TGAAGGTGTTGAAGCTGTTTACCAGGACTGGCTGAAAAAACGTCGTGCTC     850
TGCCGTTCGAAGCTGACGGTGTTGTTGTTCGTCTGGACGAACTGGCTCTG     900
TGGCGTGAACTGGGTTACACCGCTCGTGCTCCGCGTTTCGCTATCGCTTA     950
CAAATTCCCGGCTGAAGAAAAAGAAACCCGTCTGCTGGACGTTGTTTTCC     1000
AGGTTGGTCGTACCGGTCGTGTTACCCCGGTTGGTATCCTGGAACCGGTT     1050
```

TTCCTGGAAGGTTCTGAAGTTTCTCGTGTTACCCTGCACAACGAATCTTA 1100
CATCGAAGAACTGGACATCCGTATCGGTGACTGGGTTCTGGTTCACAAAG 1150
CTGGTGGTGTTATCCCGGAAGTTCTGCGTGTTCTGAAAGAACGTCGTACC 1200
GGTGAAGAACGTCCGATCCGTTGGCCGGAAACCTGCCCGGAATGCGGTCA 1250
CCGTCTGCTGAAAGAAGGTAAAGTTCACCGTTGCCCGAACCCGCTGTGCC 1300
CGGCTAAACGTTTCGAAGCTATCCGTCACTTCGCTTCTCGTAAAGCTATG 1350
GACATCCAGGGTCTGGGTGAAAAACTGATCGAACGTCTGCTGGAAAAAGG 1400
TCTGGTTAAAGACGTTGCTGACCTGTACCGTCTGCGTAAAGAAGACCTGG 1450
TTGGTCTGGAACGTATGGGTGAAAAATCTGCTCAGAACCTGCTGCGTCAG 1500
ATCGAAGAATCTAAAAAACGTGGTCTGGAACGTCTGCTGTACGCTCTGGG 1550
TCTGCCGGGTGTTGGTGAAGTTCTGGCTCGTAACCTGGCTGCTCGTTTCG 1600
GTAACATGGACCGTCTGCTGGAAGCTTCTCTGGAAGAACTGCTGGAAGTT 1650
GAAGAAGTTGGTGAACTGACCGCTCGTGCTATCCTGGAAACCCTGAAAGA 1700
CCCGGCTTTCCGTGACCTGGTTCGTCGTCTGAAAGAAGCTGGTGTTGAAA 1750
TGGAAGCTAAAGAAAAAGGTGGTGAAGCTCTGAAAGGTCTGACCTTCGTT 1800
ATCACCGGTGAACTGTCTCGTCCGCGTGAAGAAGTTAAAGCTCTGCTGCG 1850
TCGTCTGGGTGCTAAAGTTACCGACTCTGTTTCTCGTAAAACCTCTTACC 1900
TGGTTGTTGGTGAAAACCCGGGTTCTAAACTGGAAAAAGCTCGTGCTCTG 1950
GGTGTTCCGACCCTGACCGAAGAAGAACTGTACCGTCTGCTGGAAGCTCG 2000
TACCGGTAAAAAAGCTGAAGAACTGGTT

> Theta+LGT
TccggcgtagaggatcgagatcgatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattTtgtttaactttaagaaggagatatacatatgcaccatcatcatcatcattcttctggtctggtgccacgcggt
tctggtatgaaagaaaccgctgctgctAaattcgaacgccagcacatggacagcccagatctgggtaccgacgacgacgacaag

ATGACCCTGGAAGAAGCTCGTAAACGTGTTAACGAACTGCGTGACCTGATCCGTTACCACAACTACCGTTACTACGTTCTGGCTGACCCGGAAATCTCTGACGCTGAATACGACCGTCTGCTGCGTGAACTGAAAGAACTGGAAGAACGTTTCCCGGAACTGAAATCTCCGGAC
TCTCCGACCCTGCAGGTTGGTGCTCGTCCGCTGGAAGCTACCTTCCGTCCGGTTCGTCACCCGACCCGTATGTACTCTCTGGACAACGCTTTCAACCTGGACGAACTGAAAGCTTTCGAAGAACGTATCGAACGTGCTCTGGGTCGTAAAGGTCCGTTCGCTTACACCGTTGAA
CACAAAGTTGACGGTCTGTCTGTTAACCTGTACTACGAAGAAGGTGTTCTGGTTTACGGTGCTACCCGTGGTGACGGTGAAGTTGGTGAAGAAGTTACCCAGAACCTGCTGACCATCCCGACCATCCCGCGTCGTCTGAAAGGTGTTCCGGAACGTCTGGAAGTTCGTGGTGAA
GTTTACATGCCGATCGAAGCTTTCCTGCGTCTGAACGAAGAACTGGAAGAACGTGGTGAACGTATCTTCAAAAACCCGCGTAACGCTGCTGCTGGTTCTCTGCGTCAGAAAGACCCGCGTATCACCGCTAAACGTGGTCTGCGTGCTACCTTCTACGCTCTGGGTCTGGGTCTG
GAAGAAGTTGAACGTGAAGGTGTTGCTACCCAGTTCGCTCTGCTGCACTGGCTGAAAGAAAAAGGTTTCCCGGTTGAACACGGTTACGCTCGTGCTGTTGGTGCTGAAGGTGTTGAAGCTGTTTACCAGGACTGGCTGAAAAAACGTCGTGCTCTGCCGTTCGAAGCTGACGGT
GTTGTTGTTCGTCTGGACGAACTGGCTCTGTGGCGTGAACTGGGTTACACCGCTCGTGCTCCGCGTTTCGCTATCGCTTACAAATTCCCGGCTGAAGAAAAAGAAACCCGTCTGCTGGACGTTGTTTTCCAGGTTGGTCGTACCGGTCGTGTTACCCCGGTTGGTATCCTGGAA
CCGGTTTTCCTGGAAGGTTCTGAAGTTTCTCGTGTTACCCTGCACAACGAATCTTACATCGAAGAACTGGACATCCGTATCGGTGACTGGGTTCTGGTTCACAAAGCTGGTGGTGTTATCCCGGAAGTTCTGCGTGTTCTGAAAGAACGTCGTACCGGTGAAGAACGTCCGATC
CGTTGGCCGGAAACCTGCCCGGAATGCGGTCACCGTCTGCTGAAAGAAGGTAAAGTTCACCGTTGCCCGAACCCGCTGTGCCCGGCTAAACGTTTCGAAGCTATCCGTCACTTCGCTTCTCGTAAAGCTATGGACATCCAGGGTCTGGGTGAAAAACTGATCGAACGTCTGCTG
GAAAAAGGTCTGGTTAAAGACGTTGCTGACCTGTACCGTCTGCGTAAAGAAGACCTGGTTGGTCTGGAACGTATGGGTGAAAAATCTGCTCAGAACCTGCTGCGTCAGATCGAAGAATCTAAAAAACGTGGTCTGGAACGTCTGCTGTACGCTCTGGGTCTGCCGGGTGTTGGT
GAAGTTCTGGCTCGTAACCTGGCTGCTCGTTTCGGTAACATGGACCGTCTGCTGGAAGCTTCTCTGGAAGAACTGCTGGAAGTTGAAGAAGTTGGTGAACTGACCGCTCGTGCTATCCTGGAAACCCTGAAAGACCCGGCTTTCCGTGACCTGGTTCGTCGTCTGAAAGAAGCT
GGTGTTGAAATGGAAGCTAAAGAAAAAGGTGGTGAAGCTCTGAAAGGTCTGACCTTCGTTATCACCGGTGAACTGTCTCGTCCGCGTGAAGAAGTTAAAGCTCTGCTGCGTCGTCTGGGTGCTAAAGTTACCGACTCTGTTTCTCGTAAAACCTCTTACCTGGTTGTTGGTGAA
AACCCGGGTTCTAAACTGGAAAAAGCTCGTGCTCTGGGTGTTCCGACCCTGACCGAAGAAGAACTGTACCGTCTGCTGGAAGCTCGTACCGGTAAAAAAGCTGAAGAACTGGTTtag

References:
2004. Nucleic Acids Res. 32(3);1197-207. PMID: 14973201
1998. Acta Biochim Pol. 45(3);653-60. PMID:9918481
2002. Proc Natl Acad Sci USA. 99(2): 596-601. PMID: 11782527
2014. Mol Biol Int. 2014:287430. PMID:25197572
2022. ACS Catalysis. 2022, 12, XXX, 2988-2997
2002. Mol Microbial. 43(6):1505-15. PMID: 11971263
1989. Virology. 172(1):302-10. PMID:2475971
2018. Nat Biotechnology. 36(7):645-450. PMID:29912208
2002. Mol Microbial. 43(6):1505-15. PMID: 11971263
1991. Gene. 108(1):1-6. PMID: 1761218
1976. J Bacteriol. 127(3):1550-7. PMID: 8432
1977. Biochim Biophys Acta. 475(1):32-41
1999. Proc Natl Acad Science USA. 96(1):38-43. PMID: 9874768
1996. Nucleic Acids Res. 24(15):3071-8
https://openbioeconomy.org/projects/open-enzyme-collections/